UNIVERSIDAD NACIONAL DEL LITORAL

FACULTAD DE INGENIERÍA Y CIENCIAS HÍDRICAS

ESTADÍSTICA

NOTAS DE INTRODUCCIÓN A LA TEORÍA

UNIDAD 4 – MODELOS PROBABILÍSTICOS

Responsable de cátedra: Prof. Juan Pablo Taulamet

Equipo de cátedra: ***Auxiliares:*** Lic. María José Llop (JTP) - Ing. Ana Lisa Eusebi (JTP) - Prof. Fátima Bolatti (JTP) - Ing. Franco Nardi (Ay. 1°) ***Ayudantes:*** AIA Cristian Bottazzi - Téc. Eliana García

Carreras: Ingenierías en: Recursos Hídricos - Ambiental - Agrimensura

AÑO ACADÉMICO 2024

SOBRE ESTE DOCUMENTO

Las presentes notas, se encuentran destinadas a brindar una orientación para el estudio de los modelos probabilísticos que abordaremos en la unidad 4, durante el cursado de la asignatura Estadística, en el dictado para las carreras: Ingeniería en Recursos Hídricos, Ingeniería Ambiental e Ingeniería en Agrimensura.

Tal como se ha establecido en las pautas de cursado, en el contexto de la unidad 4, se abordarán algunas distribuciones particulares de variables aleatorias discretas y continuas, a saber:

- Variable discreta: Binomial, Poisson, Geométrica e Hipergeométrica.
- Variable continua: Normal, Exponencial y Gamma. De valores extremos: Gumbel y Weibull.

A su vez, en el contexto de la Unidad 6, se utilizarán algunos modelos vinculados al modelo Normal: t-Student, Chi cuadrado y F de Fisher-Snedecor. Los mismos serán abordados en un material anexo.

A continuación se pretende realizar un abordaje básico acerca de cada modelo propuesto, dejando de lado una gran cantidad de demostraciones, propiedades, relaciones y derivaciones posibles, con el fin de ofrecer una lectura fluida que permita orientar el estudio y la realización de las actividades prácticas. Para completar el estudio, se recomienda profundizar mediante la lectura de bibliografía complementaria.

Este documento fue actualizado por última vez el día 25/09/24 a las 11:09:31 hs.

ACERCA DE LOS MODELOS PROBABILÍSTICOS

En el contexto del dictado de nuestra asignatura, un modelo probabilístico es una herramienta que a partir de una teorización, nos permitirá asociar una variable aleatoria (VA) con una y/o varias funciones (Sean de cuantía o densidad, así como de distribución) de probabilidades y con un conjunto de características que le son propias a dicho modelo. Esta asociación se llevará a cabo bajo ciertos supuestos y permitirá obtener probabilidades asociadas a determinados valores de la VA. De manera inversa se podrá a partir de determinados valores de probabilidades obtener valores de la VA.

LA ASOCIACIÓN DE UNA VA CON UN MODELO PROBABILÍSTICOS

Cuando se efectúa la asociación de una VA con un modelo probabilístico, es habitual utilizar expresiones tales como: “La VA sigue la distribución tal…”, un ejemplo sería la siguiente expresión: “X sigue la distribución Binomial”. En ese contexto, se parte del supuesto de que la VA en cuestión sigue un comportamiento tal como el que propone dicho modelo, lo cual implica una asociación entre valores de dicha variable con determinados valores de probabilidad y también con determinadas características del modelo, como por ejemplo una esperanza μ o un desvío σ. De forma equivalente a lo anterior, pero en términos simbólicos, se establece la siguiente expresión: X~Binomial. A su vez, se podrán incluir determinados valores que denominaremos parámetros del modelo, los cuales se explicitarán entre paréntesis y separados por coma. De modo general, la sintaxis para la definición de la asociación es la siguiente: nombre de la variable ~ nombre del modelo (valor parámetro 1, valor parámetro 2, ...).

A continuación abordaremos los modelos discretos y continuos más importantes, estableciendo para cada caso: el contexto en el que se puede realizar la asociación de una VA con la distribución de probabilidades, los parámetros del modelo, las funciones *f*(x) y F(x) y las características más importantes. También incluiremos algunas de las relaciones que se pueden establecer entre los diferentes modelos y algunas singularidades.

MODELOS PROBABILÍSTICOS DISCRETOS

MODELO BERNOULLI

Este modelo puede utilizarse cuando se cumplen las condiciones para considerar lo que se denomina un “Ensayo de Bernoulli”. Para el planteo de un ensayo de Bernoulli, es preciso definir un evento en el EM, el cual tendrá una probabilidad de ocurrencia asociada denotada por p y será considerado un como un “Éxito”. En ese contexto, se considerará el evento complementario al Éxito, bajo el nombre de Fracaso y por probabilidad contraria se establecerá que la probabilidad de que suceda dicho fracaso será 1-p. Cabe aclarar que las definiciones del éxito y el fracaso en términos de Bernoulli se encuentran despojadas de cualquier valoración positiva y/o negativa, de manera que sería posible que un éxito se encontrare eventualmente asociado a situaciones desafortunadas y simplemente se trata de un evento, que incluye uno o más puntos del EM que son de interés para estudio que se realiza. En términos concretos, cuando se produce el éxito la VA toma el valor 1 y cuando se produce un fracaso el valor 0.

Parámetros

El único parámetro de este modelo se denomina p y representa la probabilidad de éxito del ensayo de Bernoulli, es decir la probabilidad de que la variable tome el valor 1.

Funciones

$$f(x) = \begin{cases} 1-p & \text{si } x=0 \\ p & \text{si } x=1 \end{cases}$$

$$F(x) = \begin{cases} 0 & x<0 \\ 1-p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Características más importantes

$$
\begin{aligned}
\mathrm{E}[X] &= 0 \cdot \mathrm{P}[X = 0] + 1 \cdot \mathrm{P}[X = 1] \\
&= p
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}\,[X] &= \mathrm{E}[X^2] - \mathrm{E}[X]^2 \\
&= p - p^2 \\
&= p\,(1 - p)
\end{aligned}
$$

Para el cálculo de la varianza se ha considerado:

$$
\begin{aligned}
\mathrm{E}[X^2] &= \sum_{x=0}^{1} x^2 p^x (1-p)^{1-x} \\
&= p(1-p)^{1-1} \\
&= p
\end{aligned}
$$

Contexto de aplicación

El modelo Bernoulli, puede aplicarse siempre que pueda establecerse con claridad qué es lo que se entiende por éxito del ensayo y cuál es su probabilidad de ocurrencia. Este modelo permite establecer las condiciones para la definición de los otros modelos, como ser: Binomial, Poisson, Geométrico e Hipergeométrico.

Simbología

A modo de ejemplo, si se supone que la VA denotada por X sigue la distribución de Bernoulli, con un parámetro p que toma el valor 10%, en términos simbólicos puede expresarse X~Bernoulli(10%) y en ese contexto P(X=1) representa la probabilidad de que se obtenga un éxito.

MODELO BINOMIAL

El presente modelo considera la posibilidad de realizar n repeticiones de un evento que pueda ser considerado un Ensayo de Bernoulli. Es necesario que cada una de las repeticiones del ensayo se realice en las mismas condiciones, lo cual implica que la probabilidad de éxito en cada ensayo es constante y permite asociar este modelo con una variable que cuenta la cantidad de éxitos, de tal forma que podrá tomar valores discretos desde 0 y hasta n.

Parámetros

Este modelo posee dos parámetros denotados por n y p. El parámetro n representa la cantidad de veces que se repite el experimento correspondiente al ensayo de Bernoulli. El parámetro p representa la probabilidad de éxito en cada ensayo.

Funciones

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

siendo

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Este coeficiente permite cuantificar el total de combinaciones posibles de obtener x éxitos en n pruebas

$$F(x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$$

Características más importantes

$$\mathrm{E}[X] = np(p+1-p)^{n-1} = np$$

$$\mathrm{V}[X] = np(1-p)$$

Contexto de aplicación

El modelo Binomial puede aplicarse siempre que pueda identificarse con claridad un evento que pueda ser considerado como un éxito, en el contexto de un ensayo de Bernoulli, y que pueda repetirse en las mismas condiciones un número determinado de veces.

Simbología

A modo de ejemplo, si se supone que la VA denotada por X sigue la distribución Binomial, con los parámetros n y p que toman los valores 5 y 20% respectivamente, en términos simbólicos puede expresarse X~Bin(5,20%) y en ese contexto P(X=2) representa la probabilidad de que se encuentren dos éxitos luego de repetir el ensayo 5 veces con una probabilidad de éxito del 20% cada vez.

MODELO GEOMÉTRICO

El presente modelo considera la posibilidad de realizar n repeticiones de un evento que pueda ser considerado un Ensayo de Bernoulli. Al igual que en el modelo anterior, es necesario que cada una de las repeticiones del ensayo se realice en las mismas condiciones, lo cual implica que la probabilidad de éxito en cada ensayo es constante, pero este modelo una variable con la cantidad de veces que debe repetirse el ensayo para que se obtenga un éxito por primera vez.

Parámetros

Este modelo posee un sólo parámetro denotado por p. que representa la probabilidad de éxito en términos de un ensayo de Bernoulli si el experimento se realiza una vez.

Lo que representa la variable

La variable aleatoria N que sigue una distribución Geométrica cuenta la cantidad de repeticiones necesarias hasta encontrar el primer éxito de en un ensayo de Bernoulli que posee una probabilidad de éxito p.

Funciones

$$f(n) = p(1-p)^{n-1}$$

$$F(n) = 1-(1-p)^{n-1}$$

Características más importantes

$$E(X) = \frac{1}{p}$$

$$V(X) = \frac{1-p}{p^2}$$

Contexto de aplicación

El modelo Geométrico puede aplicarse siempre que pueda identificarse con claridad un evento que pueda ser considerado como un éxito, en el contexto de un ensayo de Bernoulli, y que sea de interés analizar la cantidad de pruebas que deben realizarse para obtener un conjunto sucesivo de fracasos que culmine con un primer éxito.

Simbología

A modo de ejemplo, si se supone que la VA denotada por N sigue la distribución Geométrica, con un parámetro p = 15%, en términos simbólicos puede expresarse X~Geom(15%). En ese contexto, la expresión P(N=2) representa la probabilidad de que la segunda repetición del ensayo sea la primera exitosa.

Singularidades: Pérdida de memoria

Este modelo posee una singularidad vinculada al cálculo de las probabilidades condicionales, de tal forma que, si X es una VA que sigue la distribución Geométrica, y dados dos valores de m y b positivos, se verifica que:

$$\mathrm{P}[X > m+n | X > m] = \mathrm{P}[X > n]$$

MODELO POISSON

Para establecer una solución de continuidad con lo anterior, se propone pensar el modelo Poisson como una extensión con el modelo Binomial con parámetros n y p, en condiciones tales que n toma valores grandes (formalmente tiende a infinito) y p valores pequeños (formalmente tiende a cero). De esta forma, si se considera que la probabilidad de éxito p toma un valor cercano a cero, puede decirse que dicho éxito es un suceso raro. En este contexto, -y de manera similar al modelo anterior-, la VA que sigue la distribución Poisson, es discreta y cuenta la cantidad de éxitos, es decir la cantidad de veces que ocurre algún suceso raro en algún contexto determinado.

Parámetros

Este modelo tiene un sólo parámetro denotado por ƛ, que representa la esperanza de la variable y posee la singularidad de que también representa la varianza. En consonancia con lo planteado anteriormente, si se toma una VA Binomial con parámetros n y p, el parámetro ƛ puede calcularse como el producto de n y p.

Lo que representa la variable

La variable aleatoria X que sigue una distribución Poisson cuenta la cantidad de veces que ocurre un suceso raro en algún contexto determinado.

Funciones

$$f(x) = \frac{\lambda^x e^{-\lambda}}{n!}$$

$$F(x) = \sum_{\forall x_i \le x} f(x_i)$$

Características más importantes

$$\mathrm{E}[X] = \sum_{k=0}^{\infty} k\left(\frac{e^{-\lambda}\lambda^k}{k!}\right) = e^{-\lambda}\sum_{k=1}^{\infty}\frac{k\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} =$$

$$= \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

$$\mathrm{V}[X] = \lambda$$

Contexto de aplicación

Según se ha planteado anteriormente, el modelo Poisson puede utilizarse en el contexto de una VA que sigue una distribución Binomial(n,p), si n tiende a infinito y p tiende a 0 (La consideración de cuán grande debe ser n y cuán pequeña debe ser p, se dejará librada a los términos de interpretación subjetivos que eventualmente deberán fundamentarse). Además de lo anterior, es posible considerar que una VA discreta sigue una distribución

Poisson, si la misma cuenta la cantidad de veces que ocurre un suceso raro -es decir un evento que posee una baja probabilidad de ocurrencia-, en un período determinado de tiempo. Análogamente, esto podría pensarse en relación a contextos diferentes al tiempo, por ejemplo una distancia, un volumen u otros.

Singularidades: Estacionariedad e igualdad entre la esperanza y la varianza

Este modelo posee la singularidad de que su esperanza y su varianza alcanzan toman el mismo valor. De esta forma, cuando se pueda reconocer que para una determinada VA su esperanza y su varianza coinciden o toman valores muy cercanos, se podría suponer que dicha variable sigue una distribución Poisson.

Se considera que una VA que sigue la distribución Poisson, se vincula con el conteo de un proceso estacionario en el sentido de la proporcionalidad de la esperanza en el contexto de la misma. Nos referimos a la proporcionalidad en los términos de la regla de tres simple y se propone un ejemplo que resultará práctico para ilustrar lo anterior. Dada una VA que sigue la distribución Poisson, con la siguiente definición:

X: “Cantidad de veces que se produce una contaminación severa del aire en un período de un año”

Y suponiendo, que $E(X) = \lambda = 1.5$

Si se define una nueva variable Y, similar a X pero considerando un período de tiempo diferente, por ejemplo:

Y: “Cantidad de veces que se produce una contaminación severa del aire en un período de dos años”

Entonces, se considerará que $E(Y) = 3$.

De esta forma estamos estableciendo una proporcionalidad entre la esperanza de la variable y el período de tiempo analizado, de tal forma de considerar que si se esperan 1.5 eventos en un año, se esperan 3 eventos en dos años y por regla de tres simple se podría establecer cualquier otro período, por ejemplo para 10 años de la siguiente forma:

1 Año → 1.5 eventos

10 Años → 10 * 1.5 / 1 = 15 eventos

Lo anterior podría simbolizarse mediante acompañando el parámetro con un subíndice:

$\lambda_{1\text{ año}} = 1.5$ $\lambda_{2\text{ años}} = 3$ $\lambda_{10\text{ años}} = 15$

Simbología

A modo de ejemplo, si se supone que la VA denotada por X sigue la distribución Poisson, con una esperanza $\lambda = 15$, en términos simbólicos se puede expresar $X \sim Poi(15)$. En ese contexto, la expresión P(X=3) representa la probabilidad de que se produzcan 3 eventos en el contexto analizado.

MODELO HIPERGEOMÉTRICO

Este modelo propone una variable que representa el conteo de la cantidad de éxitos a partir de repeticiones de un ensayo de Bernoulli, pero realizando extracciones sin reposición. Esto implica que cada ensayo no se repite en las mismas condiciones y por lo tanto, la probabilidad de éxito p es diferente. De esta forma, existe una dependencia entre cada ensayo realizado y los resultados obtenidos en los ensayos anteriores, de tal forma que la probabilidad de éxito p no es constante.

Para estudiar lo antedicho, el modelo considera un conjunto de elementos definidos como una Población, de la cual se extrae un subconjunto de elementos, que conformarán lo que se considera una Muestra. En ese contexto, el modelo propone, a partir de conocer la cantidad de elementos que son compatibles con el éxito -en términos de ensayo de Bernoulli- en la Población, considerar una VA discreta, que cuenta la cantidad de elementos que son compatibles con el éxito en la muestra y sin reponer cada elemento considerado. A diferencia del modelo Binomial, la probabilidad de que se produzca un éxito será diferente en cada extracción.

Una ejemplificación sencilla de lo anterior, podría ser pensar en términos de probabilidad clásica, que si en una extracción se contase con 2 casos favorables al éxito entre 10 posibles, es decir p= 20%, le siguiente extracción no poseerá el mismo valor de p, ya que al no poseer reposición, los casos posibles serán 9 y los casos favorables podrán ser 1 o 2, dependiendo del resultado previo, es decir, p = 1/9 ≈ 11.11% o p≈22.22%.

Parámetros

Este modelo posee tres parámetros:

N: Cantidad de elementos que posee la población.

n: Cantidad de elementos que posee la muestra.

K: Cantidad de elementos compatibles con el Éxito en la población.

Lo que representa la variable

La variable aleatoria X que sigue una distribución Hipergeométrica es una VA discreta que cuenta la cantidad de elementos compatibles con el Éxito en la muestra.

Funciones

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad F(x) = \sum_{\forall x_i \le x} f(x_i)$$

Características más importantes

$$\mathrm{E}[X] = \frac{nK}{N}$$

$$\mathrm{Var}[X] = \frac{nK}{N}\left(\frac{N-K}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Contexto de aplicación

Según se ha planteado anteriormente, este modelo puede aplicarse cuando se se considera una población de elementos, de los cuales existe un conjunto que son compatibles con un éxito en términos de un ensayo de Bernoulli, pero del cual se extrae una muestra sin reposición y se desea estudiar la cantidad de elementos exitosos en la misma.

Singularidades: Dependencia entre extracciones

Este modelo posee la singularidad de que cada vez que se realiza un experimento, es decir, cada vez que produce una extracción, el elemento obtenido como resultado no se repone al conjunto de elementos posibles para la próxima extracción, lo que genera que la probabilidad de éxito no sea constante y que exista dependencia entre las extracciones. Esta dependencia contrasta con lo que proponen otros modelos, como por ejemplo el Binomial, en el cual, cada vez que se realiza un experimento, todos resultados posibles son los mismos y la probabilidad de éxito constante en cada extracción.

Otra singularidad del modelo Hipergeométrico, es que no considera la cantidad de éxitos en un conjunto de repeticiones que se realizan de una vez, sino la cantidad éxitos obtenidos en un subconjunto de elementos de la población considerados una muestra.

Simbología

A modo de ejemplo, si se supone que la VA denotada por X sigue la distribución Hipergeométrica, con parámetros N = 20, k=10 y n=5, en términos simbólicos se puede expresar X~HG(20,10,5) y en ese contexto la expresión P(X=3) representa la probabilidad de que si se extrae una muestra de 5 elementos, de una población de 20 elementos de los cuales 10 son compatibles con un éxito -en términos de Bernoulli-, se obtengan 3 elementos compatibles con el éxito en dicha muestra.

MODELOS PROBABILÍSTICOS CONTINUOS

MODELO EXPONENCIAL

Este modelo puede pensarse a partir de un escenario similar al del modelo de Poisson. Si se supone que existe una VA discreta que sigue dicha distribución, de tal forma que la misma cuenta la cantidad de eventos raros que ocurren en un período de tiempo, puede plantearse una VA exponencial que mide el tiempo hasta la ocurrencia del próximo suceso raro.

Parámetros

Este modelo tiene el mismo parámetro que el modelo Poisson, denotado por λ, que representa la inversa de la esperanza de la variable.

Lo que representa la variable

La variable aleatoria T que sigue una distribución Exponencial mide el tiempo hasta la próxima ocurrencia de un suceso raro.

Funciones

$$f(t) = \begin{cases} 0 & \text{para } t \leq 0 \\ \lambda e^{-\lambda t} & \text{para } t > 0 \end{cases} \qquad F(t) = \begin{cases} 0 & \text{para } t \leq 0 \\ 1 - e^{-\lambda t} & \text{para } t > 0 \end{cases}$$

Características más importantes

$$\mathrm{E}[T] = \frac{1}{\lambda}$$

$$\mathrm{V}[T] = \frac{1}{\lambda^2}$$

Contexto de aplicación

Según se ha planteado anteriormente, el modelo Exponencial puede utilizarse en el contexto de una VA que mide la cantidad de tiempo necesario hasta la próxima ocurrencia de un proceso de Poisson.

Singularidades: Falta de memoria

Este modelo posee una singularidad vinculada al cálculo de las probabilidades condicionales, de tal forma que, si T es una VA que sigue la distribución Exponencial, y dados dos valores de t y u positivos, se verifica que:

$$\mathrm{P}[T > t + u / T > u] = \mathrm{P}[T > t]$$

Para ilustrar lo anterior se propone un ejemplo práctico. Dada una VA que sigue la distribución Poisson, con la siguiente definición:

X: “Cantidad de veces que se produce una contaminación severa del aire en un período de un año”

Y suponiendo, que E(X) = ƛ = 1.5

Si se define una nueva variable T, que midel el tiempo hasta la ocurrencia de la próxima contaminación en las mismas unidades temporales, de la siguiente forma:

T: “Tiempo hasta que se produzca la próxima contaminación severa del aire en años”

En ese contexto, la probabilidad condicional siguiente:

$\mathrm{P}[T > 3/T > 2]$

Será equivalente a la siguiente probabilidad:

$\mathrm{P}[T > 1]$

Simbología

A modo de ejemplo, si se supone que la VA denotada por T sigue la distribución Exponencial, con una esperanza 1/ƛ = 1/15, en términos simbólicos se puede expresar T~Exp(15). En ese contexto, la expresión P(T>2) representa la probabilidad de que se produzca un evento de tipo Poisson en más de 2 unidades de tiempo.

MODELO GAMMA

Este modelo considera una VA que mide el tiempo hasta la k-ésima ocurrencia de un evento de tipo Poisson. De esta forma, el modelo Exponencial, puede pensarse como un caso especial del modelo Gamma cuando k=1. Otra forma de pensar el modelo es como la suma de un conjunto k variables exponenciales independientes con que poseen un mismo parámetro ƛ,

Parámetros

Este modelo posee dos parámetros denotados por k y ƛ, que representan la cantidad de sucesos raros que se consideran y la esperanza de la cantidad de sucesos raros por unidad de tiempo. También puede usarse una notación con parámetros α y β de manera que α considera la cantidad de sucesos raros (equivalente a k) y β el inversa de la esperanza del proceso de Poisson (equivalente a 1/ƛ).

Lo que representa la variable

La variable aleatoria T que sigue una distribución Gamma mide el tiempo hasta la k-ésima ocurrencia de un suceso raro.

Funciones

$$f(t) = \frac{\lambda}{\Gamma(k)}(\lambda t)^{k-1} e^{-\lambda t} \text{ para } t > 0 \qquad F(t) = \int_0^t \frac{\lambda}{\Gamma(k)}(\lambda t)^{k-1} e^{-\lambda t}\, dt$$

Características más importantes

$$\mathrm{E}[T] = \frac{k}{\lambda}$$

$$\mathrm{V}[T] = \frac{k}{\lambda^2}$$

Contexto de aplicación

Según se ha planteado anteriormente, el modelo Gamma puede utilizarse en el contexto de una VA que mide la cantidad de tiempo necesario hasta la k-ésima ocurrencia de un proceso de Poisson.

Simbología

A modo de ejemplo, si se supone que la VA denotada por T sigue la distribución Gamma, con parámetros k= 2 y ƛ = 3, en términos simbólicos se puede expresar T~Γ(2,3). En ese contexto, la expresión P(T>2) representa la probabilidad de que se produzcan dos eventos de tipo Poisson en más de 2 unidades de tiempo.

MODELO NORMAL

La distribución Normal, también conocida como distribución de Laplace-Gauss o distribución gaussiana es la distribución continua más importante vinculada con el estudio de la Teoría de la Probabilidad y la Estadística. El descubrimiento es atribuido a Abraham de Moivre en el año 1733. Su aparición es muy frecuentemente en el estudio de VA vinculadas con la naturaleza y a partir del Teorema Central del Límite. Su impacto en la sociedad ha sido enorme a punto tal de modificar la idea de lo que es normal, en contraposición con lo anormal bajo un sentido patológico, es por ello que suele escribirse en forma capitalizada. La función de densidad de esta VA toma una forma conocida con el nombre de campana de Gauss.

Teorema Central del Límite

El teorema central del límite indica que, en condiciones muy generales, si S_n es la suma de n variables aleatorias independientes idénticamente distribuidas, con media y varianza finitas, entonces la función de distribución de S_n se aproxima a una distribución Normal, siempre que la suma de estas VA independientes sea lo suficientemente grande.

Parámetros

Este modelo posee dos parámetros denotados por μ y σ, que representan la esperanza y el desvío típico de la VA.

Lo que representa la variable

La variable mide alguna magnitud continua.

Funciones

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}\, dt$$

Características más importantes

$\mathrm{E}[X] = \mu$

$\mathrm{V}[X] = \sigma^2$

Contexto de aplicación

Según se ha planteado anteriormente, este modelo puede aplicarse a un gran número de variables aleatorias continuas presentes en la naturaleza e incluso puede usarse para aproximar algunos modelos discretos como el Binomial y el Poisson.

Simbología

A modo de ejemplo, si se supone que la VA denotada por X sigue la distribución Normal, con parámetros μ = 30 y σ = 2, en términos simbólicos se puede expresar X~N(30,2). En ese contexto, la expresión P(X>2) representa la probabilidad de que la VA tome valores mayores a 2.

Modelo Normal Estándar

El modelo Normal estándar es el nombre que recibe un caso particular de la distribución Normal cuando μ = 0 y σ= 1 y suele representarse a la VA con la letra Z.

Proceso de estandarización de una VA Normal

Suponiendo que se cuenta con una VA X~N(30,2), se puede obtener la versión estandarizada de cualquier VA Normal, mediante la resta de la esperanza y dividiendo por la desviación típica, en símbolos:

$$Z = \frac{X-\mu}{\sigma}$$

Esta nueva VA posee las siguientes funciones:

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} \qquad F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt$$

A partir de dicha definición, la nueva VA será normal estándar, es decir Z~N(0,1).

Aunque en la actualidad es posible realizar los cálculos de las probabilidades de las VA Normales utilizando computadoras, el surgimiento de la estandarización de las VA Normales con media 0 y desvío 1 ha permitido simplificar la elaboración de tablas matemáticas con las probabilidades acumuladas de la VA con un enorme potencial de aplicación.

Suma de variables normales

A partir del teorema central del límite (TCL), puede suponerse que la suma de VA Normales independientes, también sigue una distribución normal siempre y cuando la suma sea lo suficientemente grande.

Sean $X_{1,}$ $X_{2,}$ ...X_n, un conjunto de variables independientes y distribuidas bajo la forma X_1~N(μ_1,σ_1), X_2~N(μ_2,σ_2), …, X_n~N(μ_n,σ_n), y sea S definida como:

$$S = \sum_{i=1}^{n} x_i$$

Entonces la esperanza y la varianza de S, se podrán definir como:

$$E(S) = \sum_{i=1}^{n} \mu_i$$

$$V(S) = \sum_{i=1}^{n} \sigma_i^2$$

Independientemente del TCL, cualquier VA definida como una combinación lineal de VA Normales, también será Normal.

MODELOS DE VALORES EXTREMOS

Utilizaremos este tipo distribuciones de probabilidad cuando sea preciso estudiar los máximos o valores mínimos de alguna VA. A partir de la teoría de valores extremos (VE), presenta una división de este tipo de VA en tres familias de distribuciones. Estas tre familias pueden ser asociadas una a distribución con los mismos parámetros, mediante una generalización conocida como GEV.

Distribución generalizada de valores extremos (GEV)

A partir del teorema de Fisher-Tippet-Gnedenko, que considera $\{X_1, \ldots, X_n\}$ como una sucesión de de VA independientes distribuidas de forma idéntica y $M_n = \text{máx}\{X_1, \ldots, X_n\}$, y en condiciones que fueron demostradas por Gnedenko en 1943, la función G(x) tomará la forma de alguna de las tres familias de distribuciones: Gumbel, Fréchet o Weibull.

$$G(x) = \exp\left[-\left(1+\zeta\left(\frac{x-\mu}{\beta}\right)\right)_{+}^{-\frac{1}{\zeta}}\right]$$

Parámetros

μ: localización

β: escala

ζ: forma

Acerca del parámetro de forma ζ

Para $\zeta > 0$ se tiene la distribución tipo II (Fréchet).

Para $\zeta < 0$ se tiene la distribución tipo III (Weibull).

Con el límite cuando $\zeta \to 0$ se tiene la distribución tipo I (Gumbel).

A continuación presentaremos brevemente los modelos de VE de

MODELO GUMBEL (TIPO I)

Puede utilizarse para modelar la distribución del máximo o el mínimo de una VA. La teoría de valores extremos indica que es probable que sea útil si la VA sigue una distribución Normal o Exponencial.

Parámetros

Este modelo posee dos parámetros denotados por μ y β considerados que corresponden a la ubicación y escala respectivamente. En función de la GEV, podría incluirse el parámetro $\zeta = 0$.

Lo que representa la variable

Es una VA que representa a su vez valores máximos o mínimos de una VA.

Funciones

$$f(x) = \frac{1}{\beta} * e^{-(x-\mu)/\beta} e^{-e^{-(x-\mu)/\beta}} \qquad F(x) = \exp(-e^{-(x-\mu)/\beta})$$

Características más importantes

$$\mathrm{E}[X] = \mu + \gamma\beta$$

Considerando la constante de Euler-Mascheroni como,

$$\gamma \approx 0.5772$$

$$\mathrm{V}[X] = \frac{(\beta\pi)^2}{6}$$

Considerando la constante pi como,

$$\pi \approx 3.1416$$

Contexto de aplicación

Según se ha planteado anteriormente, el modelo Gumbel puede utilizarse en el contexto de una VA que mide valores máximos o mínimos de una VA. Este modelo es especialmente utilizado para describir fenómenos tales como caudales máximos diarios en períodos de un ano, picos de descargas horarias durante crecidas anuales, caudales máximos instantáneos anuales.

Singularidades: Probabilidades de retorno

El periodo de retorno de un evento cualquiera puede entenderse como el tiempo promedio o esperado que transcurre hasta que dicho evento vuelva a ocurrir. En ese contexto, el período de retorno T(x), se puede relacionar con la probabilidad de ocurrencia según la siguiente expresión:

$$T(x) = \frac{1}{P(X > x)}$$

A modo de ejemplo, si la probabilidad de que X supere a un determinado valor en un año es del 10%, se espera que el tiempo de retorno de ese evento es 10 años.

Variante opuesta para mínimos

La expresión general del modelo de Gumbel, puede aplicarse a valores máximos, no obstante si se quiere aplicar a valores mínimos, se debe plantear una variante para mínimos, con las siguientes funciones:

$$f(x) = \frac{1}{\beta} * e^{(x-\mu)/\beta} e^{-e^{(x-\mu)/\beta}} \qquad F(x) = 1 - \exp(-e^{(x-\mu)/\beta})$$

Simbología

A modo de ejemplo, si se supone que la VA denotada por X sigue la distribución Gumbel, con parámetros μ = 0 y β = 1, en términos simbólicos se puede expresar X~Gum(0,1). En ese contexto, la expresión P(X>2) representa la probabilidad de que se produzca la VA tome un valor máximo mayor a 2.

MODELO WEIBULL (TIPO III)

Este modelo permite estudiar la fiabilidad de los sistemas, a partir de la distribución de los tiempos de falla, cuando la tasa de fallos es proporcional a una potencia del tiempo. El mismo puede pensarse como un modelo relacionado con el Exponencial.

Parámetros

Este modelo posee dos parámetros denotados por α y β, vinculados a la forma y a la escala respectivamente.

El parámetro α indica si la tasa de fallos es creciente (cuando α > 1), decreciente (cuando α < 1), o permanece constante (cuando α = 1). En este caso, la distribución de la VA sigue el modelo Exponencial, es decir, el modelo Exponencial puede pensarse como un caso particular del modelo Weibull cuando α = 1.

Lo que representa la variable

La variable aleatoria T que sigue una distribución Weibull mide el tiempo hasta que se produzca un fallo, es por ello que debe tomar valores positivos.

Funciones

$$f(t) = \beta\alpha(\beta t)^{\alpha-1}e^{-(\beta t)^{\alpha}} \qquad t > 0$$

$$F(x) = 1 - e^{-(\beta x)^{\alpha}} \qquad t > 0$$

Características más importantes

$$\mathrm{E(T)} = \frac{1}{\beta}\,\Gamma\left(1 + \frac{1}{\alpha}\right)$$

$$\mathrm{V(T)}=\frac{1}{\beta^2}\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]$$

Contexto de aplicación

Según se ha planteado anteriormente, el modelo Weibull puede utilizarse en el contexto de una VA que mide la cantidad de tiempo hasta la próxima ocurrencia de un fallo. Suele usarse en el estudio de la fiabilidad de los sistemas. Una gran mayoría de los equipos no tienen una tasa de fallos constante: es más probable que fallen a medida que envejecen. En ese caso la tasa de fallos es creciente. Aunque también es posible encontrar equipos con tasas de fallos decrecientes. El término fiabilidad se vincula con la probabilidad de buen funcionamiento de algo que puede definirse como el complemento de la función acumulativa del modelo de Weibull.

Simbología

A modo de ejemplo, si se supone que la VA denotada por T sigue la distribución Weibull, con parámetros $\alpha = 1$ y $\beta = 20$, en términos simbólicos se puede expresar T~Weibull(1,20). En ese contexto, la expresión P(T>2) representa la probabilidad de que se produzca un evento fallo en más de 2 unidades de tiempo.

OTRAS DISTRIBUCIONES

Es posible el estudio de otras distribuciones de probabilidades que no son parte de este material y se recomienda que en caso de estudiarlas se intente construir un resumen con las posibles aplicaciones, parámetros, funciones y características más importantes. Algunas de estas distribuciones son:

Modelo de Fréchet

Se trata de una distribución para valores extremos de Tipo II.

Modelo LogNormal y su vinculación con el modelo Normal

La distribución normal logarítmica se utiliza cuando se cuenta con una VA X tal que e^x~LogNormal(μ, σ). Debido a su vinculación con el modelo Normal, es posible definir una nueva variable como Y = ln(X), y será cierto la misma seguirá una distribución Normal con los mismos parámetros, es decir que Y~N(μ,σ), de tal forma que P(X<x) = P(Y< ln(x)).

Distribuciones t de Student, Chi cuadrado, F de Snedecor

Estas distribuciones serán abordadas en un material anexo, vinculado a la unidad 6, que refiere a la Estadística Inferencial.